

The Structure of the Hereditary Material

An account of the investigations which have led to the formulation of an understandable structure for DNA. The chemical reactions of this material within the nucleus govern the process of reproduction

by F. H. C. Crick

Viewed under a microscope, the process of mitosis, by which one cell divides and becomes two, is one of the most fascinating spectacles in the whole of biology. No one who watches the event unfold in speeded-up motion pictures can fail to be excited and awed. As a demonstration of the powers of dynamic organization possessed by living matter, the act of division is impressive enough, but even more stirring is the appearance of two identical sets of chromosomes where only one existed before. Here lies biology's greatest challenge: How are these fundamental bodies duplicated? Unhappily the copying process is beyond the resolving power of microscopes, but much is being learned about it in other ways.

One approach is the study of the nature and behavior of whole living cells; another is the investigation of substances extracted from them. This article will discuss only the second approach, but both are indispensable if we are ever to solve the problem; indeed some of the most exciting results are being obtained by what might loosely be described as a combination of the two methods.

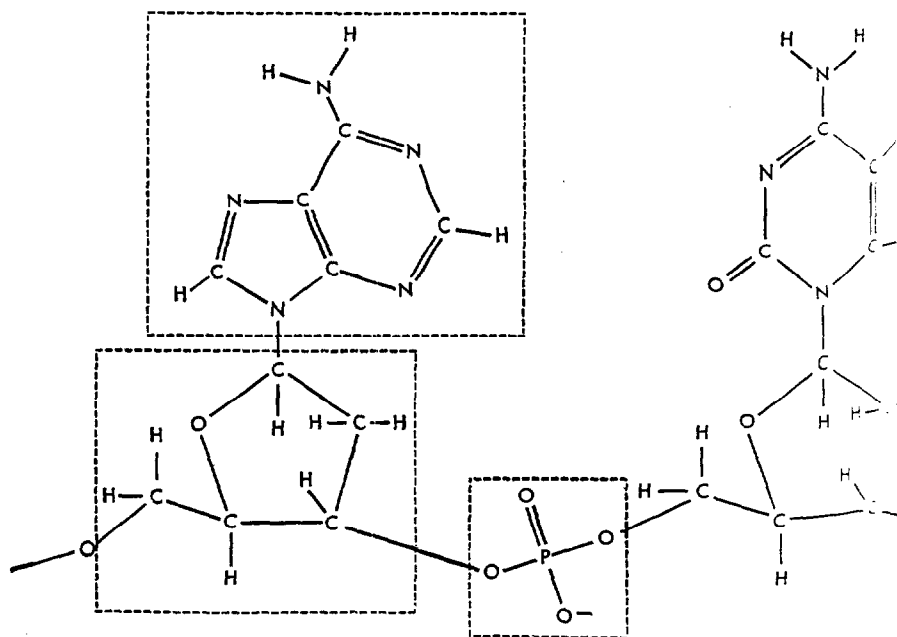
Chromosomes consist mainly of three kinds of chemical: protein, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). (Since RNA is only a minor component, we shall not consider it in detail here.) The nucleic acids and the proteins have several features in common. They are all giant molecules, and each type has the general structure of a main backbone with side groups attached. The proteins have about 20 dif-

ferent kinds of side groups; the nucleic acids usually only four (and of a different type). The smallness of these numbers itself is striking, for there is no obvious chemical reason why many more types of side groups should not occur. Another interesting feature is that no protein or nucleic acid occurs in more than one optical form; there is never an optical isomer, or mirror-image molecule. This shows that the shape of the molecules must be important.

These generalizations (with minor

exceptions) hold over the entire range of living organisms, from viruses and bacteria to plants and animals. The impression is inescapable that we are dealing with a very basic aspect of living matter, and one having far more simplicity than we would have dared to hope. It encourages us to look for simple explanations for the formation of these giant molecules.

The most important role of proteins is that of the enzymes—the machine tools of the living cell. An enzyme is specific,



FRAGMENT OF CHAIN of deoxyribonucleic acid shows the three basic units that make up the molecule. Repeated over and over in a long chain, they make it 1,000 times as long

often highly specific, for the reaction which it catalyzes. Moreover, chemical and X-ray studies suggest that the structure of each enzyme is itself rigidly determined. The side groups of a given enzyme are probably arranged in a fixed order along the polypeptide backbone. If we could discover how a cell produces the appropriate enzymes, in particular how it assembles the side groups of each enzyme in the correct order, we should have gone a long way toward explaining the simpler forms of life in terms of physics and chemistry.

We believe that this order is controlled by the chromosomes. In recent years suspicion has been growing that the key to the specificity of the chromosomes lies not in their protein but in their DNA. DNA is found in all chromosomes—and only in the chromosomes (with minor exceptions). The amount of DNA per chromosome set is in many cases a fixed quantity for a given species. The sperm, having half the chromosomes of the normal cell, has about half the amount of DNA, and tetraploid cells in the liver, having twice the normal chromosome complement, seem to have twice the amount of DNA. This constancy of the amount of DNA is what one might expect if it is truly the material that determines the hereditary pattern.

Then there is suggestive evidence in two cases that DNA alone, free of protein, may be able to carry genetic information. The first of these is the dis-

covery that the "transforming principles" of bacteria, which can produce an inherited change when added to the cell, appear to consist only of DNA. The second is the fact that during the infection of a bacterium by a bacteriophage the DNA of the phage penetrates into the bacterial cell while most of the protein, perhaps all of it, is left outside.

The Chemical Formula

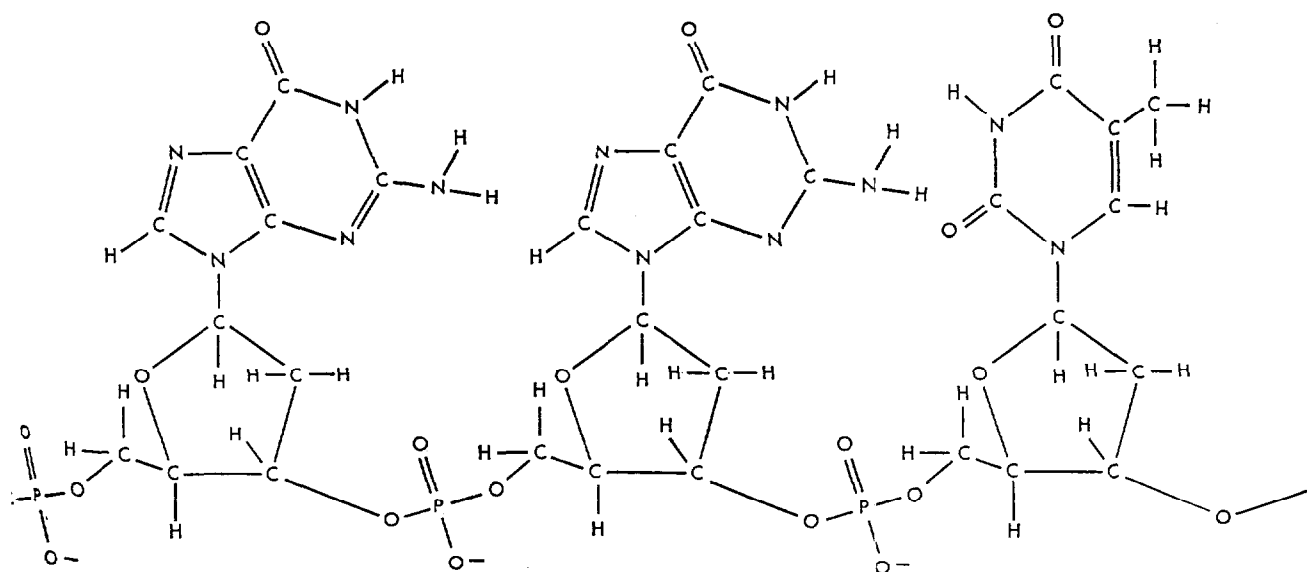
DNA can be extracted from cells by mild chemical methods, and much experimental work has been carried out to discover its chemical nature. This work has been conspicuously successful. It is now known that DNA consists of a very long chain made up of alternate sugar and phosphate groups [see diagram below]. The sugar is always the same sugar, known as deoxyribose. And it is always joined onto the phosphate in the same way, so that the long chain is perfectly regular, repeating the same phosphate-sugar sequence over and over again.

But while the phosphate-sugar chain is perfectly regular, the molecule as a whole is not, because each sugar has a "base" attached to it and the base is not always the same. Four different types of base are commonly found: two of them are purines, called adenine and guanine, and two are pyrimidines, known as thymine and cytosine. So far as is known the order in which they follow one another along the chain is irregular, and

probably varies from one piece of DNA to another. In fact, we suspect that the order of the bases is what confers specificity on a given DNA. Because the sequence of the bases is not known, one can only say that the *general* formula for DNA is established. Nevertheless this formula should be reckoned one of the major achievements of biochemistry, and it is the foundation for all the ideas described in the rest of this article.

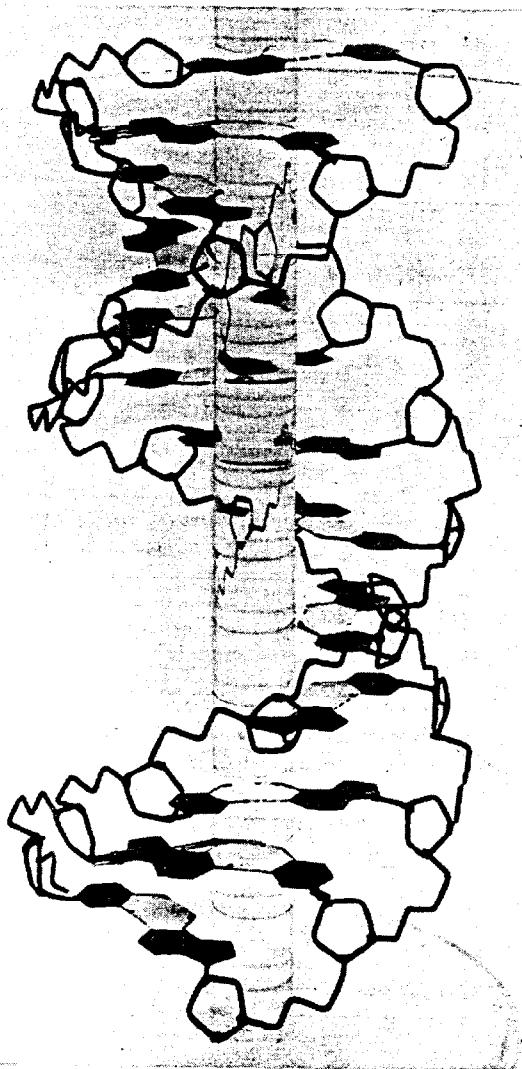
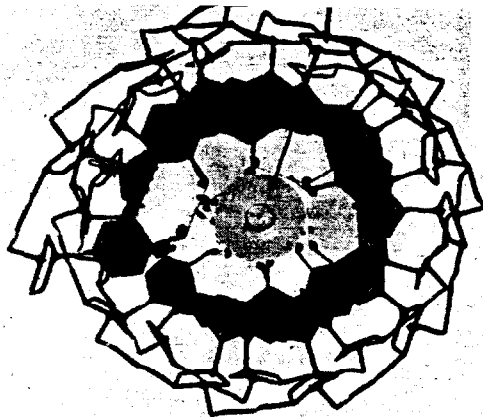
At one time it was thought that the four bases occurred in equal amounts, but in recent years this idea has been shown to be incorrect. E. Chargaff and his colleagues at Columbia University, A. E. Mirsky and his group at the Rockefeller Institute for Medical Research and G. R. Wyatt of Canada have accurately measured the amounts of the bases in many instances and have shown that the relative amounts appear to be fixed for any given species, irrespective of the individual or the organ from which the DNA was taken. The proportions usually differ for DNA from different species, but species related to one another may not differ very much.

Although we know from the chemical formula of DNA that it is a chain, this does not in itself tell us the shape of the molecule, for the chain, having many single bonds around which it may rotate, might coil up in all sorts of shapes. However, we know from physical-chemical measurements and electron-microscope pictures that the molecule usually is long, thin and fairly straight,



as it is thick. The backbone is made up of pentose sugar molecules (marked by the middle colored square), linked by phosphate

groups (bottom square). The bases (top square), adenine, cytosine, guanine and thymine protrude off each sugar in irregular order.



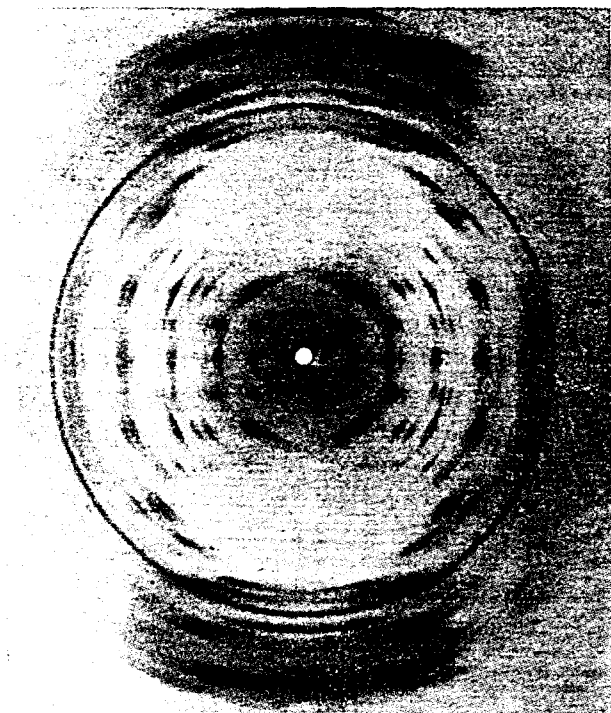
STRUCTURAL MODEL shows a pair of DNA chains wound as a helix about the fiber axis. The pentose sugars can be plainly seen. From every one on each chain protrudes a base, linked to an opposing one at the same level by a hydrogen bond. These base-to-base links act as horizontal supports, holding the chains together. Upper photograph is a top view.

rather like a stiff bit of cord. It is only about 20 Angstroms thick (one Angstrom = one 100-millionth of a centimeter). This is very small indeed, in fact not much more than a dozen atoms thick. The length of the DNA seems to depend somewhat on the method of preparation. A good sample may reach a length of 30,000 Angstroms, so that the structure is more than 1,000 times as long as it is thick. The length inside the cell may be much greater than this, because there is always the chance that the extraction process may break it up somewhat.

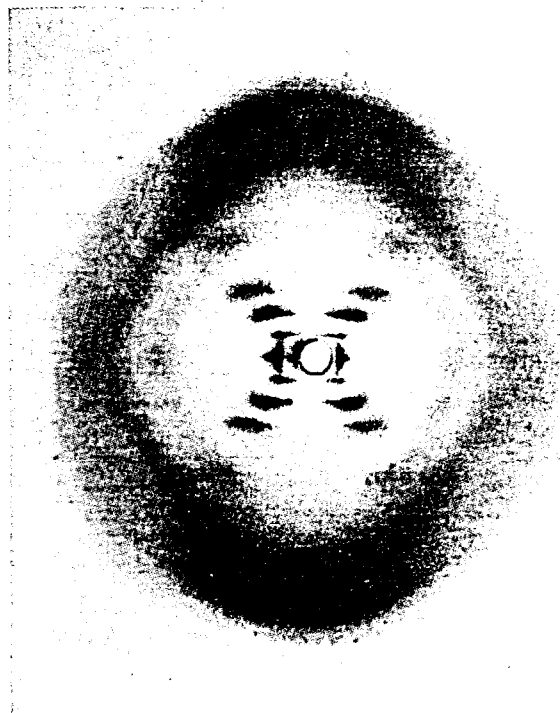
Pictures of the Molecule

None of these methods tells us anything about the detailed arrangement in space of the atoms inside the molecule. For this it is necessary to use X-ray diffraction. The average distance between bonded atoms in an organic molecule is about $1\frac{1}{2}$ Angstroms; between unbonded atoms, three to four Angstroms. X-rays have a small enough wavelength ($1\frac{1}{2}$ Angstroms) to resolve the atoms, but unfortunately an X-ray diffraction photograph is not a picture in the ordinary sense of the word. We cannot focus X-rays as we can ordinary light; hence a picture can be obtained only by round-about methods. Moreover, it can show clearly only the periodic, or regularly repeated, parts of the structure.

With patience and skill several English workers have obtained good diffraction pictures of DNA extracted from cells and drawn into long fibers. The first studies, even before details emerged, produced two surprises. First, they revealed that the DNA structure could take two forms. In relatively low humidity, when the water content of the fibers was about 40 per cent, the DNA molecules gave a crystalline pattern, showing that they were aligned regularly in all three dimensions. When the humidity was raised and the fibers took up more water, they increased in length by about 30 per cent and the pattern tended to become "paracrystalline," which means that the molecules were packed side by side in a less regular manner, as if the long molecules could slide over one another somewhat. The second surprising result was that DNA from different species appeared to give identical X-ray patterns, despite the fact that the amounts of the four bases present varied. This was particularly odd because of the existence of the crystalline form just mentioned. How could the structure appear so regular when the bases varied? It seemed that the



STRUCTURE A is the crystalline form of DNA found at relatively low humidity. This X-ray photograph is by H. R. Wilson.



STRUCTURE B is the paracrystalline form of DNA. The molecules are less regularly arranged. Picture is by R. E. Franklin.

broad arrangement of the molecule must be independent of the exact sequence of the bases, and it was therefore thought that the bases play no part in holding the structure together. As we shall see, this turned out to be wrong.

The early X-ray pictures showed a third intriguing fact: namely, that the repeats in the crystallographic pattern came at much longer intervals than the chemical repeat units in the molecule. The distance from one phosphate to the next cannot be more than about seven Angstroms, yet the crystallographic repeat came at intervals of 28 Angstroms in the crystalline form and 34 Angstroms in the paracrystalline form; that is, the chemical unit repeated several times before the structure repeated crystallographically.

J. D. Watson and I, working in the Medical Research Council Unit in the Cavendish Laboratory at Cambridge, were convinced that we could get somewhere near the DNA structure by building scale models based on the X-ray patterns obtained by M. H. F. Wilkins, Rosalind Franklin and their co-workers at Kings' College, London. A great deal is known about the exact distances between bonded atoms in molecules, about the angles between the bonds and about the size of atoms—the so-called van der Waals' distance between adjacent non-bonded atoms. This information is easy

to embody in scale models. The problem is rather like a three-dimensional jig saw puzzle with curious pieces joined together by rotatable joints (single bonds between atoms).

The Helix

To get anywhere at all we had to make some assumptions. The most important one had to do with the fact that the crystallographic repeat did not coincide with the repetition of chemical units in the chain but came at much longer intervals. A possible explanation was that all the links in the chain were the same but the X-rays were seeing every tenth link, say, from the same angle and the others from different angles. What sort of chain might produce this pattern? The answer was easy: the chain might be coiled in a helix. (A helix is often loosely called a spiral; the distinction is that a helix winds not around a cone but around a cylinder, as a winding staircase usually does.) The distance between crystallographic repeats would then correspond to the distance in the chain between one turn of the helix and the next.

We had some difficulty at first because we ignored the bases and tried to work only with the phosphate-sugar backbone. Eventually we realized that we had to take the bases into account,

and this led us quickly to a structure which we now believe to be correct in its broad outlines.

This particular model contains a pair of DNA chains wound around a common axis. The two chains are linked together by their bases. A base on one chain is joined by very weak bonds to a base at the same level on the other chain, and all the bases are paired off in this way right along the structure. In the diagram opposite, the two ribbons represent the phosphate-sugar chains, and the pairs of bases holding them together are symbolized as horizontal rods. Paradoxically, in order to make the structure as symmetrical as possible we had to have the two chains run in opposite directions; that is, the sequence of the atoms goes one way in one chain and the opposite way in the other. Thus the figure looks exactly the same whichever end is turned up.

Now we found that we could not arrange the bases any way we pleased: the four bases would fit into the structure only in certain pairs. In any pair there must always be one big one (purine) and one little one (pyrimidine). A pair of pyrimidines is too short to bridge the gap between the two chains, and a pair of purines is too big to fit into the space.

At this point we made an additional assumption. The bases can theoretically

exist in a number of forms depending upon where the hydrogen atoms are attached. We assumed that for each base one form was much more probable than all the others. The hydrogen atoms can be thought of as little knobs attached to the bases, and the way the bases fit together depends crucially upon where these knobs are. With this assumption the only possible pairs that will fit in are: adenine with thymine and guanine with cytosine.

The way these pairs are formed is shown in the diagrams on page 60. The dotted lines show the hydrogen bonds, which hold the two bases of a pair together. They are very weak bonds; their energy is not many times greater than the energy of thermal vibration at room temperature. (Hydrogen bonds are the main forces holding different water molecules together, and it is because of them that water is a liquid at room temperatures and not a gas.)

Adenine must always be paired with thymine, and guanine with cytosine; it is impossible to fit the bases together in any other combination in our model. (This pairing is likely to be so fundamental for biology that I cannot help wondering whether some day an enthusiastic scientist will christen his newborn twins Adenine and Thymine!) The model places no restriction, however, on the sequence of pairs along the structure. Any specified pair can follow any other. This is because a pair of bases is flat, and since in this model they are stacked roughly like a pile of coins, it does not matter which pair goes above which.

It is important to realize that the specific pairing of the bases is the direct result of the assumption that both phosphate-sugar chains are helical. This regularity implies that the distance from a sugar group on one chain to that on the other at the same level is always the same, no matter where one is along the chain. It follows that the bases linked

to the sugars always have the same amount of space in which to fit. It is the regularity of the phosphate-sugar chains, therefore, that is at the root of the specific pairing.

The Picture Clears

At the moment of writing, detailed interpretation of the X-ray photographs by Wilkins' group at Kings' College has not been completed, and until this has been done no structure can be considered proved. Nevertheless there are certain features of the model which are so strongly supported by the experimental evidence that it is very likely they will be embodied in the final correct structure. For instance, measurements of the density and water content of the DNA fibers, taken with evidence showing that the fibers can be extended in length, strongly suggest that there are two chains in the structural unit of DNA. Again, recent X-ray pictures have shown clearly a most striking general pattern which we can now recognize as the characteristic signature of a helical structure. In particular there are a large number of places where the diffracted intensity is zero or very small, and these occur exactly where one expects from a helix of this sort. Another feature one would expect is that the X-ray intensities should approach cylindrical symmetry, and it is now known that they do this. Recently Wilkins and his co-workers have given a brilliant analysis of the details of the X-ray pattern of the crystalline form, and have shown that they are consistent with a structure of this type, though in the crystalline form the bases are tilted away from the fiber axis instead of perpendicular, as in our model. Our construction was based on the paracrystalline form.

Many of the physical and chemical properties of DNA can now be understood in terms of this model. For example, the comparative stiffness of the

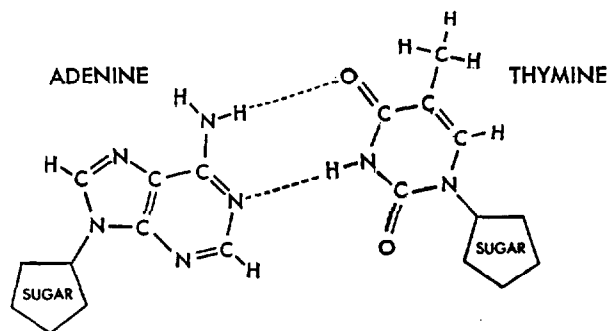
structure explains rather naturally why DNA keeps a long, fiber-like shape in solution. The hydrogen bonds of the bases account for the behavior of DNA in response to changes in pH. Most striking of all is the fact that in every kind of DNA so far examined—and over 40 have been analyzed—the amount of adenine is about equal to the amount of thymine and the guanine equal to the cytosine, while the cross-ratios (between, say, adenine and guanine) can vary considerably from species to species. This remarkable fact, first pointed out by Chargaff, is exactly what one would expect according to our model, which requires that every adenine be paired with a thymine and every guanine with a cytosine.

It may legitimately be asked whether the artificially prepared fibers of extracted DNA, on which our model is based, are really representative of intact DNA in the cell. There is every indication that they are. It is difficult to see how the very characteristic features of the model could be produced as artefacts by the extraction process. Moreover, Wilkins has shown that intact biological material, such as sperm heads and bacteriophage, gives X-ray patterns very similar to those of the extracted fibers.

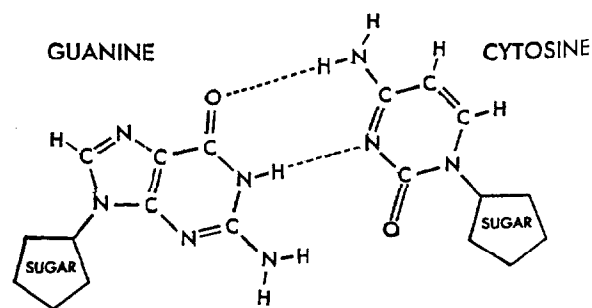
The present position, therefore, is that in all likelihood this statement about DNA can safely be made: its structure consists of two helical chains wound around a common axis and held together by hydrogen bonds between specific pairs of bases.

The Mold

Now the exciting thing about a model of this type is that it immediately suggests how the DNA might produce an exact copy of itself. The model consists of two parts, each of which is the complement of the other. Thus either chain may act as a sort of mold on which a complementary chain can be synthe-



ONE LINKAGE of base to base across the pair of DNA chains is between adenine and thymine. For the structure proposed, the link of a large base with a small one is required to fit chains together.



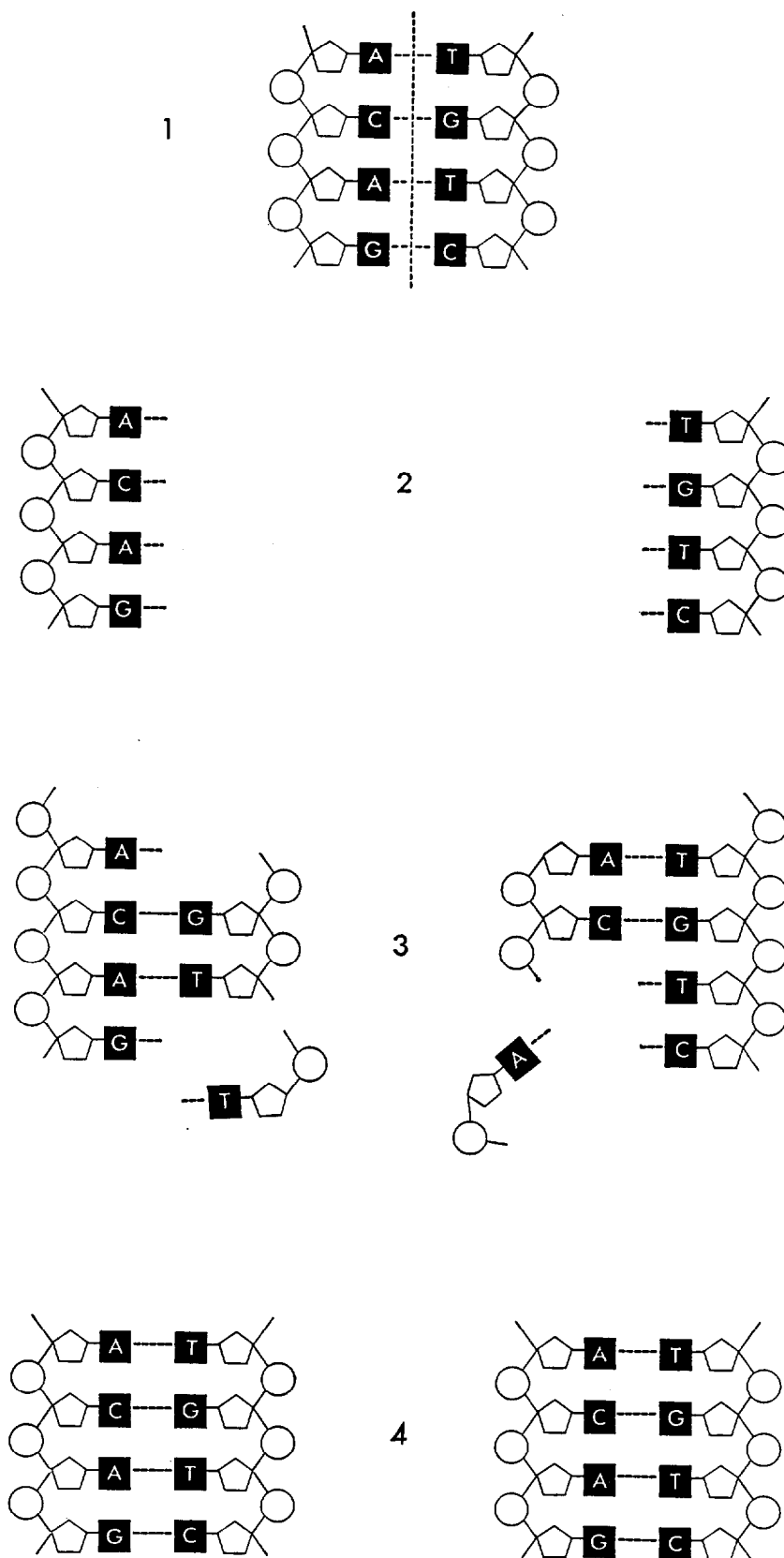
ANOTHER LINKAGE is comprised of guanine with cytosine. Assuming the existence of hydrogen bonds between the bases, these two pairings, and only these, will explain the actual configuration.

sized. The two chains of a DNA, let us say, unwind and separate. Each begins to build a new complement onto itself. When the process is completed, there are two pairs of chains where we had only one. Moreover, because of the specific pairing of the bases the sequence of the pairs of bases will have been duplicated exactly; in other words, the mold has not only assembled the building blocks but has put them together in just the right order.

Let us imagine that we have a single helical chain of DNA, and that floating around it inside the cell is a supply of precursors of the four sorts of building blocks needed to make a new chain. Unfortunately we do not know the makeup of these precursor units; they may be, but probably are not, nucleotides, consisting of one phosphate, one sugar and one base. In any case, from time to time a loose unit will attach itself by its base to one of the bases of the single DNA chain. Another loose unit may attach itself to an adjoining base on the chain. Now if one or both of the two newly attached units is not the correct mate for the one it has joined on the chain, the two newcomers will be unable to link together, because they are not the right distance apart. One or both will soon drift away, to be replaced by other units. When, however, two adjacent newcomers are the correct partners for their opposite numbers on the chain, they will be in just the right position to be linked together and begin to form a new chain. Thus only the unit with the proper base will gain a permanent hold at any given position, and eventually the right partners will fill in the vacancies all along the forming chain. While this is going on, the other single chain of the original pair also will be forming a new chain complementary to itself.

At the moment this idea must be regarded simply as a working hypothesis. Not only is there little direct evidence for it, but there are a number of obvious difficulties. For example, certain organisms contain small amounts of a fifth base, 5-methyl cytosine. So far as the model is concerned, 5-methyl cytosine fits just as well as cytosine and it may turn out that it does not matter to the organism which is used, but this has yet to be shown.

A more fundamental difficulty is to explain how the two chains of DNA are unwound in the first place. There would have to be a lot of untwisting, for the total length of all the DNA in a single chromosome is something like four centimeters (400 million Angstroms). This



REPLICATION mechanism by which DNA might duplicate itself is shown in diagram. A helix of two DNA chains unwinds and separates (1). Two complementary chains of DNA (2) within the cell begin to attach DNA precursor units floating loosely (3). When the proper bases are joined, two new helices will build up (4). Letters represent the bases.

means that there must be more than 10 million turns in all, though the DNA may not be all in one piece.

The duplicating process can be made to appear more plausible by assuming that the synthesis of the two new chains begins as soon as the two original chains start to unwind, so that only a short stretch of the chain is ever really single. In fact, we may postulate that it is the growth of the two new chains that unwinds the original pair. This is likely in terms of energy because, for every hydrogen bond that has to be broken, two new ones will be forming. Moreover, plausibility is added to the idea by the fact that the paired chain forms a rather stiff structure, so that the growing chain would tend to unwind the old pair.

The difficulty of untwisting the two chains is a topological one, and is due to the fact that they are intertwined. There would be no difficulty in "unwinding" a single helical chain, because there are so many single bonds in the chain about which rotation is possible. If in the twin structure one chain should break, the other one could easily spin around. This might relieve accumulated strain, and then the two ends of the broken chain, still being in close proximity, might be joined together again. There is even some evidence suggesting that in the process of extraction the chains of DNA may be broken in quite a number of places and that the structure nevertheless holds together by means of the hydrogen bonding, because there is never a break in both chains at the same level. Nevertheless, in spite of these tentative suggestions, the difficulty of untwisting remains a formidable one.

A Code for Heredity?

There remains the fundamental puzzle as to how DNA exerts its hereditary influence. A genetic material must carry out two jobs: duplicate itself and control the development of the rest of the cell in a specific way. We have seen how it might do the first of these, but the structure gives no obvious clue concerning how it may carry out the second. We suspect that the sequence of the bases acts as a kind of genetic code. Such an arrangement can carry an enormous amount of information. If we imagine that the pairs of bases correspond to the dots and dashes of the Morse code, there is enough DNA in a single cell of the human body to encode about 1,000 large textbooks. What we want to know, however, is just how this is done in terms of atoms and molecules. In particular,

what precisely is it a code for? As we have seen, the three key components of living matter—protein, RNA and DNA—are probably all based on the same general plan. Their backbones are regular, and the variety comes from the sequence of the side groups. It is therefore very natural to suggest that the sequence of the bases of the DNA is in some way a code for the sequence of the amino acids in the polypeptide chains of the proteins which the cell must produce. The physicist George Gamow has recently suggested in a rather abstract way how this information might be transmitted, but there are some difficulties with the actual scheme he has proposed, and so far he has not shown how the idea can be translated into precise molecular configurations.

What then, one may reasonably ask, are the virtues of the proposed model, if any? The prime virtue is that the configuration suggested is not vague but can be described in terms acceptable to a chemist. The pairing of the bases can be described rather exactly. The precise positions of the atoms of the backbone is less certain, but they can be fixed within limits, and detailed studies of the X-ray data, now in progress at Kings' College, may narrow these limits considerably. Then the structure brings together two striking pieces of evidence which at first sight seem to be unrelated—the analytical data, showing the one-to-one ratios for adenine-thymine and guanine-cytosine, and the helical nature of the X-ray pattern. These can now be seen to be two facets of the same thing. Finally, is it not perhaps a remarkable coincidence, to say the least, to find in this key material a structure of exactly the type one would need to carry out a specific replication process; namely, one showing both variety and complementarity?

The model is also attractive in its simplicity. While it is obvious that whole chromosomes have a fairly complicated structure, it is not unreasonable to hope that the molecular basis underlying them may be rather simple. If this is so, it may not prove too difficult to devise experiments to unravel it. It would, of course, help enormously if biochemists could discover the immediate precursors of DNA. If we knew the monomers from which nature makes DNA, RNA and protein, we might be able to carry out very spectacular experiments in the test tube. Be that as it may, we now have for the first time a well-defined model for DNA and for a possible replication process, and this in itself should make it easier to devise crucial experiments.